과학적 추리에 관한 인지적 확률모델 연구*

이 영 의

【주제분류】 과학철학
【주 요 어】 확률추리, 유사성, 개념 공간, 신경망 모델, 과학적 추리
【요 약 문】 이 글의 목적은 인지과학의 연구 결과를 기반으로 과학적 추리에 대한 확률적 접근을 제시하는 것이다. 이 글에서 확률 개념은 유사성의 정도로서 제안된다. 이러한 제안은 개념과 표상에 관한 최근의 인지과학의 연구 성과에 기반을 두고 있다. 인지과학자들은 그 동안의 연구를 통하여 개념들이 유사성에 의존하여 형성되고 분류된다는 점을 실험적으로 보여주고 있다. 또한 신경망 연구자들은 인간의 뇌를 모방한 인공 신경망에서 유사성에 기반을 둔 개념의 의미를 설명한다. 이러한 연구 결과들은 개념이 유사성에 기반을 두고 형성되고 학습된다는 것을 뒷받침한다. 인지과학자들은 그 동안 유사성과 확률의 관계를 거의 다루지 않았지만 최근에 들어서 그 주제에 주목하는 경향이 나타나고 있다. 나는 이 글에서 그러한 최신 동향에서 나타날 수 있는 한 가지 가능한 결과, 즉 유사성 개념에 기반을 둔 확률 개념을 도입하여 그것을 과학적 추리를 분석하는데 적용한다.

Ⅰ. 서론

과학적 추리에 관한 확률적 이론은 확률 해석에 따라서 몇 가지 유형으로 구분된다. 대표적인 확률적 이론으로는 논리적 해석에 기반을 두고 있는 카르납

* 이 논문은 2001년 한국학술진흥재단의 지원에 의하여 연구되었음.
(KRF-2001-037-BA0059)

(Carnap)의 이론, 성향적 해석에 기반을 두고 있는 포퍼(Popper)의 이론, 주관적 해석에 기반을 두고 있는 베이즈주의(Bayesianism), 그리고 빈도적 해석에 기반을 두고 있는 라이헨바하(Reichenbach)의 이론, 기어리(Giere)의 이론, 메이요(Mayo)의 오차통계학 등이 있다. 카르납, 라이헨바하, 포퍼의 이론에서 발견되는 견해 차이와 최근의 베이즈주의자들과 메이요의 논쟁에서 나타나듯이, 이러한 이론들은 확률에 관한 서로 다른 해석에 기반을 두고 있기 때문에 과학적 추리의 본성과 특징을 각자 다른 방식으로 파악한다. 확률적 접근에 관한 기존의 논의는 대체로 그러한 차이점에 초점을 맞추어 온 나머지 그것들이 공통적으로 갖고 있는 한 가지 특징을 간과해 왔다. 즉, 그 이론들은 확률적 추리의 주체인 과학자들의 인지적 측면(cognitive aspect)을 고려하지 않는다. 확률을 개인의 신념도(degree of belief)로서 간주하는 개인적 베이즈주의는 "신념"을 다루기 때문에 인지적 측면을 고려한다고 생각될 수 있지만 실제로 베이즈주의자들의 주장을 살펴보면 신념 체계를 유지하기 위한 규범적 원리와 요청에 기반을 두고 있음을 알게 된다. 이러한 현상에 대한 일반적인 비판은 규범성이 과학적 활동에서 실제로 지켜지지 않는다는 것을 지적하는 것이다. 예를 들어, 인지심리학들은 다양한 분야에 종사하는 전문가들과 과학자들을 대상으로 하는 실험을 통하여 베이즈주의가 가정하는 원리나 요청들이 경험적으로 성립하지 않는다는 점을 지적한다. 그러나 이러한 비판은 피상적일 수 있다. 보다 근본적인 문제는 베이즈주의가 요구하는 규범성이 실제로 지켜지지 않는다는 점에 있는 것이 아니라 확률을 과학적 추리에 적용하는데 있어서 인간의 인지 능력과 체계를 고려하지 않는 데에 있다. 인지적 측면에 대한 고려가 필요한가의 여부는 인지과학의 최근 성과에 비추어 보면 분명해진다. 논리적 해석, 성향적 해석, 빈도적 해석에 기반을 두고 있는 이론들도 베이즈주의와 마찬가지로 추리 담당자의 인지 체계를 고려하지 않는다는 점에서 정도의 차이는 있겠지만 동일하게 비판의 대상이 될 수 있다.

이 글의 목적은 인지과학의 연구 결과를 기반으로 과학적 추리에 대한 확률적 접근의 방향을 제시하는 것이다. 과학적 추리에 관한 인지적 확률 모델을 제안하기 위하여 이 글은 두 가지 전제로부터 출발한다. 첫째, 과학적 추리에 대한 확률적 접근은 경험적 증거와 부합해야 한다. 둘째, 확률적 모델의 경험적 증거는 인지과학이 제공하는 실험 결과와 모델들로부터 확보될 수 있다. 이러한 가정들의 정당성은 이 글에서 다루어지지 않는다. 그 대신 그러한 가정 하에서 개념과 표상에 관한 논의에서 중요한 역할을 차지하는 유사성(similarity)의 개

념에 기반을 둔 확률 개념이 제시된다. 인지과학자들은 개념과 범주에 관한 연구를 통하여 개념들이 유사성에 의하여 형성되고 분류된다는 점을 경험적으로 보여주고 있다. 한편 신경망 연구자들은 인간의 뇌를 모방한 인공 신경망(artificial neural network, 이하 신경망)에서 개념들이 유사성에 기반을 두고 구현될 수 방식을 구체적으로 보여준다. 개념들이 유사성에 기반을 두고 형성되고 이용된다는 것은 인지심리학과 신경망 분야에서의 연구를 통하여 경험적으로 뒷받침되고 있다. 그런데 한 가지 흥미로운 사실은 유사성을 연구하는 인지과학자들이 유사성과 확률의 관계에 대해서는 거의 연구를 하지 않고 있다는 점이다. 예를 들어, 신경망 이론을 심리철학과 과학철학의 분야에 적용하는 처칠랜드의(P. M. Churchland)는 유사성이 신경망에서 표현되고 개념들이 의미를 갖게 되는 방식을 논의하면서도 유사성과 확률의 관계는 다루지 않고 있다. 나는 이 글에서 인지과학의 최신 동향에서 나타날 수 있는 한 가지 가능한 결과, 즉 유사성 개념에 기반을 둔 확률 개념을 도입하여 그 개념을 과학적 추리를 분석하는데 이용한다.

논의 순서는 다음과 같다. 2절에서는 유사성에 기반을 둔 확률 개념을 제시하는데 필요한 유사성과 개념 공간이 도입된다. 3절에서는 그 개념들을 이용하여 확률이 정의되고, 유사성의 정도로서 정의된 확률이 신경망에서 구현되는 방식이 제시된다. 마지막으로 4절에서는 신경망 모델이 과학적 추리를 분석하는데 이용되는 방식을 보이기 위해서 과학적 설명과 입증에 관한 분석이 제시된다.

Ⅱ. 유사성과 개념 공간

인지과학 분야에서 유사성은 개념과 범주를 중심으로 연구되어 왔다. 개념과 범주는 철학, 언어학, 심리학 분야에서 핵심 주제에 속한다. 철학자들은 개념이 사고와 믿음 체계를 구성하는 기본 요소라고 간주해왔다. 철학적 전통에 따르면, 심성 상태는 내용을 가지며 그 내용은 개념들로 구성된다. 그렇다면 개념이란 무엇인가? 심리학자들이 고전적 견해(classical view)라고 부르는 입장에 따르면, 개념은 그것의 본질적 성질들에 의해서 정의된다. 특정 대상은 특정 개념을 정의하는 성질들을 갖고 있을 때 그 개념의 한 사례로서 분류된다. 여기서 개념은 그러한 성질에 의해서 정의될 수 있다고 가정되고 있다. 고전적 견해는 1970년대 이전에 자연종(natural kind)에 대한 철학자들의 논의에서 주류

적 경향이었으며 많은 심리학자들이 지지했던 이론이었다. 그러나 고전적 견해에 대한 많은 논의를 통하여 자연종에 대한 필요충분조건으로서의 정의를 제시할 수 없다는 것이 드러났다. 예를 들어, **얼룩말**이라는 개념을 살펴보자. **얼룩말**은 "배에 줄무늬가 있다", "초식동물이다"라는 성질을 포함한다. 그러나 배에 있는 줄무늬가 벗겨지고 인위적인 수술에 의해서 더 이상 풀을 먹을 수 없게 된 얼룩말도 여전히 **얼룩말**의 한 사례이다. 엄밀하게 말하면 "배에 줄무늬가 있다", "초식동물이다"라는 성질은 더 이상 **얼룩말**을 정의하는 성질이 될 수 없다.

이러한 문제점 때문에 심리학자들은 점차로 고전적 견해로부터 유사성에 기반을 둔 확률적 견해(probabilistic view)로 관심을 돌리게 되었다. 확률적 견해는 로쉬(Rosch)의 전형 이론(prototype theory)으로부터 비롯되었다.[1] 전형 이론은 개념에 대한 정의를 추구하는 대신 그것의 성질들을 갖는 사례들의 집합인 전형을 이용한다. 대상은 특정 개념의 전형과 충분한 유사성을 가질 때 그 개념의 사례로서 분류된다. 물론 전형 자체는 고전적 견해에서 가정되는 필요충분조건들에 의해서 정의된다. 유사성의 정도는 공유된 성질들과 보다 현저한 특징들에 의해 결정된다. 예를 들어, **새**는 "날개는 갖는다", "난다", "씨를 먹는다"라는 성질들을 포함하고, 그 중 "날개를 갖는다"라는 성질이 가장 현저한 성질이 된다. 전형 이론에서는 고전적 견해의 문제점인 자연종에 대한 정의를 제시하는 문제가 발생하지 않는다. 또한 전형 이론은 사례의 전형성을 측정할 수 있으므로 개념과 밀접한 관련이 있는 학습, 기억, 추리 현상 등을 설명하는데 경험적으로 이용될 수 있다. 예를 들어, **새**의 경우에서 전형성은 "개똥지빠귀, 갈매기, 제비, …, 플라밍고, 신천옹, 펭귄"의 순서로 낮아진다. 따라서 전형적 새에 속하는 개똥지빠귀나 갈매기는 비전형적 새에 해당하는 플라밍고나 신천옹에 비하여 보다 신속하게 효율적으로 **새**로 분류되고 학습될 수 있다. 전형적 새는 비전형적 새에 비하여 효율적으로 "X는 날 수 있다"와 같은 귀납적 판단을 가능케 한다.

전형에의 유사성에 주목하여 개념을 설명하려는 로쉬의 전략은 비트겐쉬타인

1) 확률적 견해는 로쉬의 전형 이론으로부터 표본 이론(exemplar theory)을 거쳐서 최근에는 이론-이론(theory-theory)으로 전개되고 있다. 이러한 이론들에 관한 논의는 K. Lamberts and D. Shanks(1997) *Knowledge, Concepts, and Categories*(MIT Press)와 E. Smith(1989) "Concepts and Induction" in M. Posner (ed.) *Foundations of Cognitive Science*(MIT Press), pp.501-526 참조.

(Wittgenstein)이 주장했던 언어게임 이론(language game theory)의 심리학적 버전이다. 비트겐쉬타인에 따르면 언어 행위는 카드게임과 같은 게임의 일종이다. 그는 게임과 마찬가지로 언어에서도 정의적 성질이 아니라 전체적 유사성만을 발견할 수 있다고 강조했다.[2] 개념에 관한 필요충분조건을 찾을 수 없으므로 개념들 사이에 성립하는 가족유사성(family resemblance)에 주목해야 한다는 것이다. 로쉬는 자신의 전형 이론이 가족유사성 개념을 경험적으로 지지한다고 보았다.[3] 한편, 콰인(Quine)은 논리적으로 유사성이나 자연종을 정의하는 것은 불가능하다고 지적했다.[4] 콰인의 주장이 옳다면, 유사성 개념은 논리학이나 집합론을 이용하여 정의될 수 없고 과학적으로 밝혀져야 한다. 인지심리학자들은 우리의 사고와 추리를 구성하는 개념들이 유사성을 기반으로 하고 있다는 점을 보여주었다. 우리는 유사성을 설명하기 위해서 과학적 결과를 이용해야 하는데 현재로서는 그 개념에 관한 과학적 설명을 제공할 수 있는 가장 유력한 후보는 인지과학이다.

전형 이론은 개념 분류와 학습에서 유사성이 근본적 역할을 한다는 점을 보여준다. 로쉬를 비롯한 대부분의 인지심리학자들은 개념과 전형의 관계에서 성립되는 유사성이 표현되는 공간에 대해서는 관심을 두지 않는다. 그러나 유사성을 과학적으로 설명하기 위해서는 그러한 관계가 성립하는 공간이 충분히 고려되어야 한다. 유사성이 표현되는 공간을 개념 공간(conceptual space)이라고 하자.[5] 우리의 믿음과 지식은 개념 공간에서 표현된다. 철학자들은 전통적으로 믿음과 추리를 논의하면서 추상적인 심성표상(mental representation)을 가정해왔다. 표상의 구조와 내용에 대해서는 다양한 철학 이론들이 있지만 일반적으로 표상 자체는 언어적 또는 논리적 구조를 갖는다고 생각되어 왔다. 마음에 대한 표상 이론에 따르면 명제 p에 대한 명제태도는 p를 의미하는 심성표상에 대한 믿음 관계를 갖는 것이다. 철학자들은 표상과 명제태도에 의해서 인간의

2) L. Wittgenstein(1953) *Philosophical Investigations*(Blackwell), pp.312-320.

3) E. Rosch and C. Jervis(1975) "Family Resemblance: Studies in the Internal Structure of Categories," *Cognitive Psychology* 7, p.603.

4) W. V. Quine(1969) "Natural Kinds," in *Ontological Relativity and Other Essays*(Columbia University Press), p.121.

5) 이 글에서 사용되는 "개념 공간"이라는 개념의 역사적 선례로서는 Quine(1960)의 "quality spaces," Carnap(1971)의 "attribute spaces," Stalnaker(1979)의 "logical spaces"이고 직접적으로 관련되는 것은 Gärdenfors(1990)의 "conceptual space"와 P. M. Churchland(1989)의 "state space"이다.

심리 상태와 과정을 설명한다. 그러나 항상 정당한 예외는 있다. 처칠랜드와 스티치(Stich)와 같은 제거론자들은 명제태도에 대한 가정은 과학적 근거가 없으므로 제거되어야 한다고 주장한다. 명제태도를 둘러싼 철학적 논쟁은 인지 체계를 설명하는데 있어서 적절한 분석의 차원에 대한 논쟁으로 해석될 수 있다. 적절한 분석의 차원에 대해 처칠랜드는 마(Marr)가 제시한 세 가지 분석의 차원에서 하드웨어 구현의 차원, 스티치는 알고리즘의 차원, 대부분의 철학자들은 계산적 차원을 주장한다. 그러나 인지 체계에 대한 적절한 분석의 차원은 다음에서 나타나듯이 공간 자체의 본성과 구조에 관한 논의에 대해서는 중립적이다.

인간의 뇌가 구성하는 개념 공간은 인지 내용을 충분히 표현할 정도로 다양한 차원으로 구성되어 있을 것이다. 이는 곧 개념 공간이 수많은 차원들로 구성된다는 것을 의미한다. 개념 공간은 물리학의 개념들이 표현되는 3차원(x축, y축, z축)뿐만 아니라 색이나 맛과 같은 지각적 성질을 표현하는 지각적 차원, 시간, 온도, 무게, 질량, 부피 등과 같은 물리적 성질을 표현하는 차원을 갖는다. 개념 공간은 또한 "기쁘다", "멋있다"와 같은 추상적 성질들을 표현하는 차원을 갖는다. 이러한 차원들을 고려하면 우리는 상대론적 물리적 공간을 4차원 $<x, y, z, t>$로 표현하듯이 개념 공간은 n차원 $<D_1, D_2, \cdots, D_n>$으로 표현할 수 있다. 이처럼 개념 공간을 구성하는 차원들은 발생론적으로 비언어적이다. 게르덴포스(Gärdenfors)가 지적하듯이, 인간은 행동을 계획할 때 자신의 생각들이 표현되는 언어를 가정하지 않고서도 관련된 대상들의 성질들을 생각할 수 있다.[6] 개념 공간의 차원이 추상적인 과정을 통하여 구성되는 것과는 달리 개념 공간의 구조는 과학적으로 설명될 수 있다. 신경생리학자들의 연구에 따르면 개념 공간은 위상 구조와 계량 구조를 갖고 있다. 개념 공간이 표준적인 3차원적 유클리드 공간과는 매우 다른 위상 구조와 계량 구조를 갖는다는 점이 많은 인지 실험들을 통하여 드러나고 있다. 여기서 우리는 개념 공간이 갖는 복잡성에 주목해야 한다. 개념 공간의 복잡성은 추상적 성질들과 관련된 차원의 복잡성, 위상 구조와 관련된 복잡성, 계량 구조와 관련된 복잡성으로 분류된다. 위에서 제시된 성질들의 목록에서 개념 공간은 추상적 성질의 차원을 제외하더라도 적어도 13가지의 차원을 갖고 있으며 각각의 차원에서 복잡성이 나타난다. 예를 들어, 인간의 후각은 적어도 6가지의 서로 구별되는 수용기들을 갖고 있으므로 후각의 차원은 13^6이라는 엄청난 수의 차원을 갖고 있다.[7]

---

6) P. Gärdenfors(1990) "Induction, Conceptual Spaces and AI," *Philosophy of Science* 57, pp.83-87.

이처럼 개념은 다차원적이고 비유클리드적인 위상 구조와 계량 구조를 갖는 개념 공간에서 표현된다. 전형으로서의 개념과 그 사례들은 그러한 개념 공간에서 표현된다(유사성이 개념 공간에서 표현되는 방식은 4절에서 다루어진다). 개념 공간의 위상 구조와 계량 구조에 관한 더 이상의 설명은 이 글의 목적 상 중요하지 않다. 중요한 것은 개념은 그것의 성질들의 차원들이 구성하는 개념 공간에서 표현된다는 점이다. 물리학자들이 물리적 대상의 위치와 운동을 벡터 공간을 이용하여 설명하듯이 우리는 특정한 개념을 벡터 개념을 이용하여 개념 공간상의 한 점으로 표현할 수 있다. 특정 개념이 갖는 성질들은 상대적으로 개념 공간상의 한 점으로 기술될 수 있으므로 개념 공간의 벡터는 하나의 가능한 개념을 표현한다.

Ⅲ. 신경망에서의 유사성과 확률

유사성 개념을 이용하여 확률을 규정하기 위하여 우선 유사성의 정도를 계량적으로 표현하고 그러한 결과를 이용하여 확률을 정의하기로 한다. 나는 여기서 유사성을 정의하는 대표적인 방식인 공간 모델을 이용한다.[8] 앞에서 논의되었듯이 개념들은 개념 공간상의 한 점을 나타내는 벡터에 해당한다. 두 가지 개념이 개념 공간에서 벡터 AB와 $A'B'$으로 표현된다고 가정해 보자. 락소와 코트렐(Laakso and Cottrell)이 개발한 GPA 측도(Gutman point measure)에 따르면 두 벡터의 유사성 S는 다음과 같이 정의된다.[9]

$$① \quad S(AB, A'B') = 1 - \text{평균}\left[\frac{|AB - c(A'B')|}{(AB + c(A'B'))}\right]$$

---

7) 처칠랜드의 계산에 따르면 뇌는 지각세포를 제외하고서라도 약 10^{11}개의 신경세포를 갖고 있고, 개개의 신경세포가 10^{8}개의 차원을 갖기 때문에 개념 공간의 차원 수는 10^{19}이다(P. M. Churchland, 1989, p.209),

8) 유사성을 표현하는 또 다른 방식으로는 특징에 기초를 둔 모델(feature-based model)이 있다. 이 글에서 공간 모델이 채택된 이유는 그것이 신경망에서 확률을 구현하는 방식을 설명하는데 보다 더 편리하기 때문이다. 두 가지 모델의 장단점에 대한 비교는 K. Lamberts and D. Shanks(1997), pp.54-63 참조.

9) P. M. Churchland(1998), p.19.

위의 공식에서 나타나는 상수 c는 두 벡터가 절대값은 같지만 방향이 다른 경우에 S의 값이 크게 낮아지는 경우를 교정하기 위한 교정 상수이다. 여기서 $c = \Sigma(AB)/\Sigma(A'B')$. 공식 ①은 개념 공간에서의 임의의 벡터사이의 거리를 나타내고 그 거리는 유사성의 정도를 표현한다. 예를 들어, 벡터 AB와 $A'B'$이 동일하면 유사성 S의 값은 1이 되고 그렇지 않을 경우에는 0과 1사이의 값이 나온다.

신경망은 기본적으로 두 가지 측면에서 인간의 뇌를 모의한다. 뇌는 약 10억 개 이상의 신경세포가 병렬적으로 연결되어 있는데, 이러한 뇌의 구조를 모의하기 위해서 신경망은 뇌를 구성하는 신경세포를 기본 단위(unit)로 모의하고, 신경세포의 수상돌기와 축색돌기는 전선, 시냅스는 가중치를 갖는 저항으로 모의한다. 또한 기능적 측면에서 신경망은 패턴 인식, 학습, 귀납과 같은 인지 작용을 모의하기 위해서 신경세포간의 억제와 흥분을 모의하는 가중치를 조정한다. 가중치의 수정에 의하여 나타나는 지식의 변화를 학습이라고 한다. 이러한 연결 방식을 가진 신경망의 전체 상태는 하나의 벡터로 표현된다.[10] 이러한 방식으로 유사성은 신경망에서 일관적인 방식으로 표현될 수 있다. 그러나 개념들이 표현되는 개념 공간을 성공적으로 구성하더라도 그것의 의미를 해석하는 문제는 남아있다. 어떻게 신경망에 기반을 둔 개념 공간이 의미론적으로 해석될 수 있는가? 우리는 위에서 신경망이 구현하는 개념 공간은 개념들을 표현하는 벡터들로 구성된다는 것을 보았다. 이것은 2절에서 논의된 개념의 성질과 유사성 관계는 신경망에서 구현된 개념 공간에서 점 또는 벡터로 표현된다는 것을 의미한다. 예를 들어, 전형으로서의 **새**는 벡터 AB로 표현되고, **새**의 한 사례인 개똥지빠귀는 벡터 $A'B'$로 표현된다. 그러므로 개똥지빠귀가 전형적인 새인가라는 질문은 벡터 AB와 $A'B'$의 공간적 근접성을 나타내는 식 ①을 통하여 유사성의 정도로 대답될 수 있다.

이제 처칠랜드의 신경망에 대한 의미론적 해석을 살펴보자. 처칠랜드는 신경망에서 개념들과 그것들의 유사성 관계가 표현되는 방식을 중심으로 개념 공간의 의미론을 구성하고 그것을 상태공간 의미론(state space semantics)이라고

10) 신경망은 심리철학이나 과학철학적으로 중요한 함축을 갖는다. 무엇보다도 신경망은 학습가능하기 때문에 "경험으로부터의 학습 과정"을 모의할 수 있다는 점, 인지는 기호주의가 전제하듯이 규칙 준수적이 아니라는 점을 보여준다. 신경망 모델이 갖는 본격적인 철학적 논의는 W. Bechtel and A. Abrahamsen(1991) *Connectionism and the mind*(Blackwell) 참조.

부른다. 상태공간 의미론은 하나의 형식 체계로서의 신경망 이론에 대한 의미론에 해당한다.[11] 인지 이론으로서의 신경망 이론은 상태공간 의미론이 제시됨으로써 완전한 이론 체계를 갖추었다. 그러나 신경망 이론에 비판적 입장을 취하는 포도(Fodor)는 상태공간 의미론에 대해서도 그 부적절성을 지적한다. 포도가 제시한 비판의 요지는 어떠한 두 사람도 정확히 그들의 감각 신경세포에서 동일한 인과적 의존성을 갖지 않는다는 것이다. 포도에 따르면, 사람들은 동일한 개념 공간을 가질 수 없다.[12] 포도의 비판이 옳다면, 사람들에게 공통적인 개념 공간은 존재할 수 없고 그 결과 "동일한 개념들은 각자의 개념 공간에서의 동일한 위치를 차지한다"는 처칠랜드의 입장도 성립될 수 없다. 이러한 비판에 대해, 처칠랜드는 개념 공간에서의 특정한 점이 의미론적 내용을 갖는 것은 개념 공간의 축(차원)들에 상대적인 위치의 함수로서가 아니라 개념 공간에 있는 모든 다른 점들에 대한 공간적 위치의 함수이며 외적 환경이 제공하는 안정적이고 객관적인 거시적 성질들에 대한 인과적 관계의 함수이기 때문이라고 대답한다.[13] 처칠랜드가 제시하는 답변의 요지는 유사성은 개념 공간사이의 변환에 대해서 불변적이라는 것이다. 처칠랜드는 식 ①을 이용하여 신경망이 구현하는 개념 공간에서의 유사성을 논하는데 있어서 중요한 것은 포도가 주장하듯이 개념 공간에서의 위치가 아니라 개념 공간의 내적 구조이며, 그러한 내적 구조에서 규정된 유사성은 공간 변환에 대해서 불변적이라는 점을 보여주었다. 처칠랜드는 적어도 공간 변환에 따른 의미의 변화에 대한 포도의 비판에 대해서는 성공적으로 답변한 것처럼 보인다. 하나의 개념 공간에서 표현된 유사성이 다른 공간으로 의미의 변화가 없이 변환될 수 있기 때문에 포도가 제기한 비판은 처칠랜드의 상태공간 의미론에 심각한 타격을 줄 수 없다.[14]

지금까지 우리는 개념들의 유사성을 측정하는 공식과 유사성이 신경망에서

---

11) 신경망 이론에 대한 또 다른 유력한 의미론은 P. Smolensky(1987) "The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn". *Southern Journal of Philosophy*, Supplement 26, pp.137-161 참조.

12) J. Fodor and E. Lepore(1992) *Holism: A Shopper's Guide*(Blackwell), pp.197-202.

13) P. M. Churchland(1998) "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered," *Journal of Philosophy* 98, p.8.

14) 상태공간 의미론에 대한 처칠랜드와 포도의 논쟁은 J. Fodor and E. Lepore(1992)와 R. McCauley ed(1996). *The Churchlands and Their Critics*(Blackwell) 참조.

구현되는 방식을 검토했다. 이제 남은 문제는 그러한 결과를 이용하여 확률을 정의하는 것이다. 여기서 우리는 잠시 처칠랜드가 *GPA* 측도를 이용하여 유사성을 설명하면서도 왜 그 개념을 확률에 적용하지 않았는가를 생각해 볼 필요가 있다. 유사성 개념에 기초하여 확률 개념을 정의하는 방식에 비판적인 사람들은 처칠랜드가 그러한 시도를 하지 않은 것은 그것이 개념적으로 매우 어렵거나 불가능하기 때문이라고 생각할 것이다. 그러나 처칠랜드가 그러한 시도를 하지 않은 이유는 그러한 작업이 어렵거나 불가능해서가 아니라 과학적 추리에 관한 그의 입장 때문이다. 처칠랜드는 과학적 추리는 본질적으로 최상의 설명에로의 추리(inference to the best explanation)라고 생각한다. 그는 과학적 추리에 관한 확률적 모델이 갖는 장점을 인정하지 않기 때문에 유사성을 이용하여 확률을 규정할 필요가 없다. 그러나 처칠랜드의 이러한 생각은 잘못이다. 과학적 추리를 최상의 설명에로의 추리로서 보더라도 반드시 확률적 모델을 배제할 필요는 없다. 확률적 모델은 이 글에서 시도되고 있듯이 최상의 설명에로의 추리의 틀 안에서 적용될 수 있다.

이제 유사성의 정도를 이용하여 확률을 정의하기 위해서 다음과 같은 상황을 가정해 보자. 특정한 증세 t를 가진 환자가 있는데, 그 증세는 병 A, B와 관련된다고 알려져 있다. 우리의 관심사는 증세 t를 가진 환자가 병 A에 걸렸을 확률을 구하는데 필요한 확률 개념을 제시하는 것이다. 유사성 정도 S와 확률 P는 다음과 같은 방식으로 관련된다.

② 확률(증세 t를 가진 환자가 병 A에 걸렸다) = 함수(증세 t가 병 A의 전형인 P_A와 B의 전형인 P_B에 대해 갖는 유사성).

위에서 주어진 확률과 함수의 관계는 다음의 공식으로 표현된다.

③ $P(A \mid t) = f[S(t, P_A, P_B)]$.

식 ③에서 $S(t, P_A, P_B)$는 식 ①에 의해 이미 양적으로 정의되었으므로 우리는 $f[S(t, P_A, P_B)]$에 대해 인지과학자들이 제시하는 경험적 자료를 이용하여 $P(A \mid t)$를 다음과 같이 정의할 수 있다.

④ $P(A|\ t) = \dfrac{S(t, P_A)}{S(t, P_A) + dS(t, P_B)}$ 15)

위의 식에서 나타난 상수 d는 t가 B의 전형인 P_B에 대해 갖는 유사성을 교정하는 상수이다. 상수 d가 필요한 이유는 전형 P_A에 대해 갖는 유사성의 정도에 의해서 자동적으로 전형 P_B에 대해 갖는 유사성의 정도가 결정되는 경우를 배제할 필요가 있기 때문이다. 상수 d는 "$S(t, P_A) + S(t, P_B) \neq 1$"을 보증한다. 식 ④에서 정의된 확률 $P(A \mid t)$는 조건부 확률(conditional probability)이다.

식 ④의 의미를 분명히 하기 위해서 다음 표와 같이 증세 t와 병 A, B의 유사성이 "매우 높다(H)", "거의 없다(L)"의 경우로 나누어서 검토해보자.

	$S(t, P_A)$	$S(t, P_B)$
1	H	H
2	H	L
3	L	H
4	L	L

위의 표에서 (1)의 경우 확률적 추리를 분석하는데 있어서 중요한 요소는 d의 값을 결정하는 것이다. 증세 t가 병 A, B의 전형 P_A, P_B와 동시에 유사성을 갖기 때문에 d의 값을 결정하기 위해서는 임상 실험이나 검사와 같은 경험적 증거가 필요하게 될 것이다. 예를 들어, 경험적 증거에 의해서 d의 값이 거의 0임이 알려졌다고 하자. 이 경우에 $S(t, P_A)$는 거의 1이 될 것이고, $dS(t, P_B)$는 거의 0이다. 이 경우에 $P(A \mid t)$는 거의 1에 가깝다. 즉, $P(A \mid t) \fallingdotseq 1$. (2)-(4)의 경우도 마찬가지로 분석될 수 있다.

이제 남은 문제는 조건부 확률과 구별되는 근원 사건의 확률, 즉 근원적 확

15) 공식 ③은 P. Justin, H. Nilsson, and H. Olsson, "Where Do Probability Judgments Come from? Evidence for Similarity-Graded Probability" (Unpublished)에서 제시된 공식 $P(A \mid t) = \dfrac{0.5d + S(t, P_A)}{d + S(t, P_A) + S(t, P_B)}$을 수정한 것이다.

률을 정의하는 것이다. 우리의 경우 근원적 확률은 $P(A)$와 $P(t)$이다. 근원적 확률을 정의하기 위해서는 A와 그 무엇과의 유사성 정도가 필요하게 되는데, 신경망 모델에서 그것은 신경망의 전체 활성화 상태를 표현하는 전체벡터 N에 해당한다. 그러한 유사성을 $S(N, A)$이라고 하면 $P(A) = f[S(N, A)]$이다. 공식 ①에서 제시된 유사성 정도를 이용하면 근원적 확률은 전체 벡터 N에 상대적으로 다음과 같이 정의된다.

$$⑤\quad P(A)=1-\text{평균}\left[\frac{|N-cA|}{(N+cA)}\right]$$

지금까지 논의를 통하여 우리는 근원적 확률과 조건부 확률을 유사성을 기반으로 정의할 수 있음을 보았다. 이러한 확률 개념을 과학적 추리를 분석하는데 적용하기 위해서는 확률 계산법(calculus of probability)이 필요하다. 여기서 우리는 기존의 확률 계산법과는 다른 별도의 계산법을 개발할 수 있을 것이다.[16] 그러나 위에서 정의된 근원적 확률과 조건적 확률은 기존의 확률 계산법을 이용하여 계산가능하기 때문에 독자적인 확률 계산법을 개발할 필요는 없다. 이러한 전략은 방법론적으로 독자적 확률 계산법을 개발하는 것에 비하여 다음과 같은 유리한 점을 갖고 있다. 우리는 이미 조건부 확률과 근원적 확률을 정의했기 때문에 기존의 계산법을 이용하여 확률들을 계산할 수 있다. 이는 곧 유사성에 기반을 둔 확률 개념이 기존의 확률 체계에 포섭될 수 있다는 점을 의미한다. 다른 한편으로 독자적인 확률 계산법을 개발하는 것은 유사성에 기반을 둔 확률 개념의 적용가능성을 제한하는 부정적인 결과를 낳게 될 것이다.

Ⅳ. 과학적 추리

이제 2절과 3절에서 제시된 신경망 이론과 유사성 개념에 기반을 둔 확률 개념을 이용하여 과학적 추리를 분석해보자. 우리는 그러한 개념들을 이용하여 과학적 추리를 분석하는데 있어서 나타나는 핵심 개념들을 검토함으로써 그 적용가능성과 유용성을 알게 된다.

---

16) 독자적 확률 계산법에 대한 예는 S. Block. D. Medin, and D. Osherson(2002) "Probability from Similarity"(Unpublished)를 참조

이 글의 기본 입장인 과학적 추리에 대한 인지적 접근의 관점에서 보았을 때 과학 이론의 구조는 공리적 입장이 아니라 의미론적 입장에서 파악된다. 과학 이론에 대한 공리적 입장에 따르면 이론은 형식적으로 공리화된 문장들의 집합으로서 일종의 해석된 형식 체계에 해당한다. 그러한 형식 체계에 등장하는 비논리적 용어들은 대응 원리(correspondence principle)나 의미 공준(meaning postulate)에 의해서 의미가 부여됨으로써 해석된다. 반면에 기어리와 반 프라센(van Fraassen)이 주장하는 의미론적 접근에 따르면 이론은 체계에 대한 모델의 집합과 그 모델을 실제 세계의 체계들에 관련시키는 가설로 이루어져 있다. 예를 들어, 뉴턴의 중력 이론은 중력에 대한 모델과 "태양계는 뉴턴적 중력계이다"라는 가설을 포함한다.[17] 공리적 입장과 의미론적 입장의 차이 중 현재의 논의와 관련하여 중요한 것은 다음과 같다. 즉, 전자의 관점에서 보았을 경우 전체 과학은 하나의 형식 체계로서 그 체계에 포함된 개별 이론들은 논리적 관계를 이루고 있는데 비하여, 후자의 관점에서는 과학은 모델들과 가설들의 집합이며 모델간의 관계는 유사성(기어리의 경우) 또는 경험적 적합성(반 프라센의 경우)이라는데 있다. 의미론적 입장에 따르면 과학 이론들은 연역적으로 연결되어 있는 것이 아니라 콰인(Quine)이 말했듯이 비연역적인 "신념의 그물망"(web of belief)을 형성한다.

과학 이론들이 비연역적인 망을 형성한다는 주장은 다소 애매하지만 신경망 이론을 적용하면 보다 분명하게 그 의미가 드러난다. 그 대표적인 예로서 처칠랜드가 제시한 해석을 살펴보자. 처칠랜드는 과학 이론의 구조에 관하여 기본적으로 의미론적 입장을 취하지만 전적으로 그것에 공감하지는 않는다. 처칠랜드는 무엇보다도 우리가 특정 이론을 내재화할 때 세계를 지각하는 방식을 중요시한다.[18] 처칠랜드는 2절에서 논의되었듯이 개념 또는 범주와 관련된 지각과 학습은 전형에 대한 유사성을 중심으로 이루어진다고 주장한다. 처칠랜드의 입장이 심리학적 설명과 구별되는 점은 그의 설명이 신경망에서 구현되는 벡터 공간을 중심으로 전개된다는 것이다. 특정인의 뇌에서 구현된 벡터 공간을 분할

17) R. N. Giere(1988) *Explaining Science: A Cognitive Approach*(University of Chicago Press), p.84. B. van Fraassen(1987) "The Semantic Approach to Scientific Theories," in N. Nersessian (ed.), *The Process of Science* (Kluwer), pp.105-24.

18) P. M. Churchland(1989) *A Neurocomputatinal Perspective*(MIT Press), p.158.

하는 모든 가능한 가중치들의 공간을 가정해보자. 이 경우 개념과 이론은 다음과 같이 규정된다.[19]

(1) 개념은 가중치 공간의 분할이다. 즉, 특정한 입력 벡터들이 특정한 출력 벡터를 산출토록 만드는 가능한 가중치들의 영역이다.
(2) 이론은 그러한 가중치들을 할당하는 역할을 담당한다. 특히 특정인이 갖는 전체 이론은 그의 신경망에서의 모든 가중치들을 할당한다.

(1)과 (2)가 의미하는 것은 특정인이 세계에 대해 갖는 전체 이론은 그의 신경망을 최적화 상태로 유도하는 모든 가중치들의 할당 방식이라는 점이다. 우리는 주로 교육과 학습을 통하여 세계에 관한 이론을 형성한다. 그러한 형성 과정에서 이론을 구성하는 개념들 또는 국소 이론들의 충돌을 예상할 수 있는데 우리의 인지 체계는 적절한 가중치를 부여하는 방식으로 그러한 충돌을 해결한다. 이처럼 인지 체계에서 나타나는 중요한 충돌들이 해결된 상태를 최적화 상태(state of optima)라고 하고, 개인의 전체 이론은 바로 그러한 최적화 상태를 유도하는 가중치들을 할당하는 방식이다.

우리는 2절에서 개념이나 전형은 개념 공간상의 한 점에 해당한다는 것을 보았다. 처칠랜드는 이론을 매우 구조화된 전형으로 간주하기 때문에 이론은 벡터공간상에서 특정한 점에 해당한다. 처칠랜드에 따르면, 특정인의 세계에 관한 전체 이론은 벡터공간에서 특정한 위치를 점유하는 점이며, 이론을 구성하는 개념들은 벡터공간의 분할이다.[20] 그렇다면 전체 이론과 국소 이론간의 관계는 무엇인가? 처칠랜드는 전체 이론과 국소 이론의 관계를 벡터들의 합 관계로 설명한다. 전체 이론은 학습이 벡터공간에 걸쳐서 산출되는 분할들의 집합으로 구성되며, 열역학과 같은 특정 이론은 그러한 전체 활성화 공간에서의 특정한 전형의 하위 부분으로 이해된다. 하위 부분으로서의 국소 이론은 다시 개념들이라는 하위 구조를 갖게 된다.[21] 전체 이론은 벡터공간에서 나타나는 다양한 벡터들의 합벡터이고, 전체 이론을 표현하는 합벡터를 분할하는 다양한 벡터들은 국소 이론에 해당한다. 그렇다면 어떻게 국소 이론이 공유되는가? 이 질문과 관련된 처칠랜드의 예를 들어보자.[22] 그가 제시한 예는 파동 이론을 공유했으면

19) *Ibid.*, pp.177-178.
20) *Ibid.*, p.178.
21) *Ibid.*, pp.232-234.

서도 그것의 참에 대해서는 서로 상반된 입장을 취했던 호이겐스(Huygens)와 뉴턴(Newton)에 관한 것이다. 잘 알려져 있듯이, 빛의 본성에 대하여 호이겐스는 파동 이론이 옳다고 생각하고 뉴턴은 입자 이론이 옳다고 보았다. 빛의 본성에 대해 두 과학자는 상반된 견해를 지녔다. 처칠랜드는 그들이 각자의 벡터 공간에 걸쳐서 적절하게 비슷한 분할 집합을 갖기 때문에 파동 이론을 공유하고 있었다고 주장한다. 즉, 호이겐스와 뉴턴은 파동 이론에 대한 유사한 전형을 소유하고 있으며, 그 전형은 활성화를 통하여 적절한 방식으로 비슷한 인지적, 언어적, 조작적 행위를 산출한다. 그런데 그들의 전형 이론은 상이한 방식으로 활성화된다. 호이겐스의 전형은 수면파, 음파, 광파에 접했을 때 활성화되지만, 뉴턴의 전형은 수면파와 음파에 접했을 경우에만 활성화된다. 뉴턴이 광파적 현상에 접했을 때 활성화되는 전형은 입자 이론이다. 처칠랜드의 주장에서 핵심은 "호이겐스와 뉴턴의 활성화 공간은 매우 비슷하게 분할되어 있다"는 주장이다. 어떻게 매우 다른 방식으로 가중치가 부여된 두 가지 벡터공간이 개략적으로 동일한 또는 유사한 분할을 생성할 수 있는가? 그 질문에 대한 처칠랜드의 대답은 전체 이론과 국소 이론의 구분에서 발견된다. 호이겐스와 뉴턴의 경우에서 파동 이론은 전체 이론이 아니라 국소 이론이다. 국소 이론으로서의 파동 이론은 파동 현상들에 관한 전형이다. 처칠랜드가 주장하는 것은 개똥지빠귀와 갈매기가 유사성을 중심으로 새의 사례로 분류되듯이 파동은 호이겐스와 뉴턴의 경우에서 매우 비슷한 적용 사례를 가지며 그러한 유사성은 개념 공간에서 유사한 분할을 의미한다.

지금까지 우리는 과학적 지식이 신경망에서 표현되는 방식에 대한 처칠랜드의 설명을 살펴보았다. 이제 신경망에서 표현된 지식 체계에서 발생하는 과학적 추리를 검토해 보기로 한다. 우선 유사성에 기반을 둔 확률 개념은 설명(explanation)에 적용될 수 있다. 처칠랜드에 따르면, 특정한 현상을 설명하는 것은 그 현상을 재인하는 과정과 본질적으로 동일하다. 과학적 설명은 유기체가 어떤 상황(피설명항)을 유리한 방향으로 처리하도록 만드는 전형(설명항)을 활성화하는 것이다.[23] 처칠랜드에게 있어서 어떤 대상을 지각적으로 재인하는 것과 어떤 사실을 과학적으로 설명하는 것은 모두 전형의 활성화이다. 여기서 우리는 "유기체가 어떤 상황을 유리한 방향으로 처리하도록"이라는 구절에서 처

22) P. M. Churchland(1992) "Reconceiving Cognition" in R. Giere (ed.) *Cognitive Models of Science*(University of Minnesota Press), pp.477-478.

23) *Ibid.*, p.210.

칠랜드는 과학적 설명의 본질을 최상의 설명에로의 추리라고 간주하고 있음을 알 수 있다. 그러나 우리는 지금까지 제시된 신경망 해석이 과학적 설명을 반드시 그러한 추리로 파악하게 하는 것은 아니라는 점에 주목해야 한다. 여기서 나는 신경망 이론을 이용한 과학적 추리에 관한 처칠랜드의 분석, 특히 과학적 설명에 관한 분석에 대한 비판을 제시하지는 않을 것이다.[24] 앞에서 지적되었듯이 과학적 설명에 대한 그의 분석은 최상의 설명에로의 추리라는 그의 입장을 도입하지 않고서도 이 글에서 가정되는 인지적 입장과 양립 가능하다. 과학적 설명에 대한 처칠랜드의 설명에 따르면, 설명적 이해는 잘 훈련된 신경망에서 특정한 전형의 활성화에 해당한다. 이것이 과학적 설명 또는 설명을 통한 이해의 본질을 잘 드러낸다고 인정하더라도 신경망을 이용한 신경생리학적 수준의 설명이 아니라 보다 상위 차원에서의 설명이 필요하다. 3절에서 도입된 유사성에 기반을 둔 확률 개념은 처칠랜드의 설명에서 발견할 수 없는 이론적 요구를 충족시켜줄 수 있다. 또한 피설명항을 확률적으로 설명해야 하는 경우 처칠랜드의 입장은 적용 불가능하다. 예를 들어 "어떤 방사능원소는 반감기 동안 자연 붕괴할 확률이 50%이다", "이 주사위가 3의 눈이 나올 확률은 1/6이다"와 같은 확률적 추리를 설명하기 위해서는 유사성의 정도가 아니라 확률 개념 자체가 요구된다.

유사성에 기반을 둔 확률 개념은 입증(confirmation)에도 적용될 수 있다. 특정 가설(H)을 지지하는 경험적 증거(E)가 있다고 가정해보자. 입증 관계가 어떻게 신경망 이론과 유사성에 기반을 둔 확률 개념에 의해서 분석될 수 있는가? 그 질문에 대해서 우리는 앞에서 제시된 확률 개념을 이용하여 다음과 같이 대답할 수 있다.

(3) 증거 E가 가설 H를 입증한다는 것은 E를 표현하는 벡터가 H를 표현하는 전형 벡터에 추가됨으로써 나타나는 신경망의 전체 상태는 그 이전의 상태보다 더 안정적인 최적화 상태로 이행한다는 것이다.

---

24) 처칠랜드의 입장에 대한 전반적 비판은 C. Glymour(1992), "Invasion of the Mind Snatchers" in R. N. Giere (ed.) *Cognitive Models of Science* (University of Minnesota Press), pp.465-467 참조. 처칠랜드의 접근에 대한 지지자의 입장에서 제기된 비판에 대해서는 W. Bechtel and A. Abrahamsen(1997) *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*(Blackwell), pp.292-293 참조.

물론 그러한 최적화 상태가 전체적 최적화 상태(완전한 입증)인지 아니면 국소적 최적화 상태(부분적 입증)인지는 이후에 제시되는 증거들이 그 이상의 최적화를 유도하는가에 달려있다. 여기서 중요한 점은 최적화 상태로의 이행이 유사성의 증가로 해석된다는 것이다. 즉, 가설과 증거의 유사성이 증가할수록 입증도는 증가한다. 가설과 증거의 유사성이라는 개념은 전통적인 입증 개념과는 매우 다르게 보일 것이다. 그러나 우리가 유사성을 단지 지각적으로 해석하지 않고 이론적 측면을 고려한다면 그러한 문제는 발생하지 않는다. 앞에서 지적되었듯이 이론은 모델과 해석으로 구성된다. 입증 관계에서 고려되는 유사성은 이론의 해석 부분과 직접적으로 관련된다. 해당 이론이 제시하는 모델과 실재가 제공하는 증거가 일치하는 정도를 표현하는 것이 유사성이다. 이러한 유사성 개념은 이미 과학철학자들에서 의해서 수용되고 있다. 예를 들어 기어리의 구성적 실재론(constructive realism)은 이 글에서 논의되는 유사성에 기반을 두고 있다. 기어리에 따르면, 과학자들은 이론적 모델들을 구성하면서 그것들이 실제 세계의 체계들에 대한 적어도 부분적 표현이기를 원한다. 그러므로 모델들과 세계와의 관계는 참 또는 대응 관계가 아니라 유사성의 관계이다.[25]

우리는 앞에서 모델과 세계와의 유사성의 정도를 표현할 수 있는 공식을 제시했었다. 3절에서 제시된 식 ④를 가설과 증거간의 관계를 분석하는데 적합하도록 다음과 같이 표현하기로 한다(H와 경쟁 가설이 하나인 경우만을 고려한다).

$$⑥\quad P(H|E) = \frac{S(E, H)}{S(E, H) + dS(E, -H)}$$

이제 남은 문제는 입증 관계를 위의 식을 이용하여 분석하는 것이다. 특정한 경험적 증거가 특정한 이론을 입증한다는 것은 이론과 증거간의 유사성이 증가한다는 것을 의미한다. 여기서 유사성이 증가한다는 것은 곧 확률이 증가한다는 것을 의미한다. 어떠한 확률이 증가하는가? 증거가 제시되기 전의 개념 공간에서의 해당 이론이 갖는 확률보다 증거가 제시된 이후에 그 이론이 갖는 확률이 증가한다. 이러한 결과는 흥미롭게도 베이즈주의의 입증 개념과 매우 비슷하다. 베이즈주의의 입증 개념에 따르면, 증거 E가 이론 H를 입증한다는 것은 다음의 경우이다.

---

25) R. N. Giere(1988) *Explaining Science: A Cognitive Approach*(University of Chicago Press), pp.92-93.

⑦ $P(H/E) > P(H)$

베이즈주의자들은 자신들의 입증 개념이 앞에서 제시된 방식으로 정당화될 수 있다고 생각할 수도 있다. 그러나 이러한 전략이 성공하기 위해서는 해결되어야 할 한 가지 문제점이 있다. 베이즈주의자들이 입증을 설명하기 위해서 사용하는 조건부 확률은 이 글에서 제시된 조건부 확률과 근본적으로 차이가 난다. 베이즈주의자들이 조건부 확률을 계산하는데 사용하는 다음의 베이즈 정리(Bayes's theorem)를 식 ⑥과 비교해보면 그 차이점은 분명해진다.

⑧ $$P(H/E) = \frac{P(E/H)P(H)}{P(E/H) + P(E/-H)}$$

나는 4절에서 유사성에 기반을 둔 확률 개념이 어떠한 방식으로 과학적 추리를 분석하는데 이용될 수 있는가를 과학 이론의 구조, 설명, 입증의 경우를 중심으로 논의하였다. 위에서 언급되었듯이 나의 분석은 과학적 추리를 최상의 설명에로의 추리라고 보는 관점을 제외하면 기본적으로 처칠랜드의 이론과 일치하며 또한 부분적으로 기어리의 입장과 일치한다. 이러한 일치는 우연적인 것은 아니며, 처칠랜드와 기어리는 잘 알려진 바대로 과학에 대한 인지과학적 접근을 취하는 대표적인 철학자들이다. 그러한 일치는 과학적 추리를 설명하는데 있어서 동일한 노선을 취하는 데에서 발생하는 것으로 이해되어야 한다. 그러나 입증 관계에 대한 베이즈주의적 개념과의 일치에서 볼 수 있듯이 유사성에 기반을 둔 확률 개념은 어떤 특정한 입장을 전제로 하는 것은 아니기 때문에 과학적 추리를 새로운 관점에서 분석하는데 이용될 수 있다.

V. 비판적 논의 및 결론

지금까지의 논의는 다음과 같은 세 가지 주장을 중심으로 전개되었다.

(1) 과학적 추리는 유사성에 기반을 개념화 또는 범주화를 기반으로 한다.
(2) 그러한 유사성은 뇌의 작용을 모의하는 신경망에서의 벡터공간에서 표현된다.

(3) 벡터공간에서의 유사성은 확률로서 변환될 수 있다.

(1)-(3)을 중심으로 본문에서 제시된 접근에 대해 제기될 수 있는 비판 중의 하나는 나의 분석은 유사성의 정도를 단순히 확률 개념에 연결했을 뿐이라는 지적이다. 그러나 우리는 유사성의 정도를 확률 개념에 연결함으로써 유사성의 개념만으로는 분석할 수 없는 과학적 추리를 분석할 수 있다. 나의 접근으로부터 나타나는 가장 큰 장점은 과학적 추리를 분석하는데 있어서 인지과학적 틀을 이용한 확률 계산이 가능하다는 것이다. 확률을 유사성과 관련시킴으로써 우리는 인지과학적 기반에서 확률적 추리를 분석할 수 있게 된다.

예상되는 비판 중 두 번째 부류는 제시된 확률 개념과 기존의 확률 해석간의 관계가 규명되지 않았다는 것이다. 물론 그러한 관계에 대한 후속 연구가 필요하다. 지금까지의 논의를 바탕으로 보았을 때 유사성의 정도로서의 확률 개념은 기존의 확률 해석들과 양립 가능하다. 우선 유사성은 객관적 세계를 반영하는 것이기 때문에 유사성에 기반을 둔 확률 개념은 객관적 해석으로 분류될 수 있다. 다른 한편으로 유사성의 정도로서의 확률 개념을 보다 더 기초적인 개념으로 간주하고 기존의 확률 개념들을 그러한 기초적 확률에 관한 표현이라고 볼 수도 있을 것이다. 예를 들어, 확률을 믿음의 정도로서 파악하는 주관적 해석에 대해 유사성에 기반을 둔 확률 개념은 그러한 믿음의 근거를 제시하는데 이용될 수 있다.

마지막으로 본문에서 제시된 내용은 새로운 확률 개념을 제시하는데 있어서 충분하지 못하다는 비판이 가능하다. 이러한 비판은 정당하지만 부분적으로만 그렇다. 이 글에서 제시된 접근은 아직까지 충분한 연구가 이루어지지 않았으며 시작 단계에 있다. 과학철학의 분야에서는 처칠랜드를 제외하고는 그러한 접근은 체계적으로 시도되지 않고 있으며, 처칠랜드 자신도 확률 개념의 필요성을 제기하지 않고 있다. 이러한 점에서 이 글은 앞으로의 연구를 위한 방향을 제시하는 성격을 갖는다.

(고려대 철학연구소 연구조교수)